WHAT IS CLAIMED IS:

1. A method of generating an index for a sequence that supports a non-contiguous subsequence match, comprising:

receiving a sequence;

receiving a window size;

encoding the sequence into a weighted-sequence;

encoding the weighted sequence into one or more one-dimensional sequences, wherein the length of each of the one or more one-dimensional sequences is less than the window size;

inserting each of the one or more one-dimensional sequences into a trie structure; and

generating the index, comprising:

generating a current sequential ID and a maximum sequential ID pair for

generating each of the one or more trie nodes, wherein the current sequential ID of any descendant of a given trie node is between the current sequential ID of the given trie node and the maximum sequential ID;

generating an iso-depth link for each unique symbol in each of the one or more one-dimensional sequences, wherein the iso-depth link comprises trie nodes under the symbol; and

generating an offset list comprising an original position of each of the one or more subsequences in the weighted-sequence.

2. The method of claim 1, wherein encoding the sequence into a weighted-sequence comprises encoding the sequence with weights represented by real numbers;

3. The method of claim 2, wherein encoding the sequence with weights represented by real numbers, comprises discretizing the sequence into a number of equi-width units.

4. The method of claim 1, wherein inserting each of the one or more one-dimensional sequences into a trie structure comprises using a depth-first traversal.

5. The method of claim 1, wherein creating the weighted-sequences index, wherein the weighted-sequences index comprises an iso-depth index, comprises creating the weighted-sequences index, wherein the weighted-sequences index comprises an iso-depth index, wherein the iso-depth index is a one-dimensional buffer.

6. The method of claim 1, wherein creating the weighted-sequences index, wherein the weighted-sequences index comprises an iso-depth index, comprises creating the weighted-sequences index, wherein the weighted-sequences index comprises an iso-depth index, wherein the iso-depth index is a B^+ tree.

7. The method of claim 1, wherein creating the weighted-sequences index, wherein the weighted-sequences index comprises an iso-depth index, comprises creating the weighted-sequences index, wherein the weighted-sequences index comprises an iso-depth index, wherein the iso-depth index is a linked list.

8. The method of claim 1, wherein receiving a sequence comprises receiving one or more elements in the sequence, wherein each of the one or more elements are represented by one or more (symbol, weight) pairs.

9. The method of claim 8, wherein receiving one or more elements in the sequence, wherein each of the one or more elements are represented by one or more (symbol, weight) pairs, and wherein each of the symbol elements of the one or more (symbol, weight) pairs correspond to a non-uniform frequency distribution.

10. The method of claim 9, further comprising reordering the one or more one one-dimensional sequences prior to inserting each of the one or more one-dimensional sequences into a trie structure using the non-uniform frequency distribution to generate a new sequence.

11. The method of claim 10, wherein reordering the one or more one one-dimensional sequences prior to inserting each of the one or more one-dimensional sequences into a trie structure using the non-uniform frequency distribution to generate one or more new sequences, comprises:

(a) adding an offset 2*w*r to each weight element of the one or more one-dimensional sequences, wherein w is a window size, r is a rank a symbol to generate a new weight;

(b) sorting the each element of the one or more one-dimensional sequences by the new weight;

(c) placing a moving window of size 2*w*A on the one or more new sequences, wherein A is the total number of the symbols; and

(d) indexing the one or more new sequences in a new window.

12. The method of claim 1, wherein receiving a sequence comprises receiving one or more scientific datasets, transforming each of the one or more scientific datasets into one or more sequence, concatenating the one or more sequences to form a long sequence.

13. A method of matching a query sequence in a weighted-sequences index, comprising:

receiving the query sequence;

transforming the query sequence into a weighted sequence;

encoding the weighted sequence into one or more one-dimensional sequences; and searching one or more iso-depth links of the weighted-sequences index structure using the one or more one-dimensional sequences and returning an offset.

14. The method of claim 12, wherein searching one or more iso-depth links of the weighted-sequences index structure using the one or more one-dimensional sequences and returning an offset, comprises:

```
(a) assuming the query sequence is <q_1, q_2, ..., q_n> ;
(b) assigning i=1, begin=0, end=infinity ;
(c) finding iso-depth link for q_i ;
(d) for each label pair (x, y) in the link of q_i such that
    begin<x<end do:
        (e) if (i=n) then {
                returning the offset in an offset list of nodes in
            a range of [x,y] };
        (f) if (i<n) then {
                assigning i=i+1; begin=x, end=y;
                going to step (c) }.
```

15. A machine-readable medium having instructions stored thereon for execution by a processor to perform a method of generating an index for a sequence that supports a non-contiguous subsequence match, comprising the steps of:

receiving a sequence;

receiving a window size;

encoding the sequence into a weighted-sequence;

encoding the weighted sequence into one or more one-dimensional sequences, wherein the length of each of the one or more one-dimensional sequences is less than the window size;

inserting each of the one or more one-dimensional sequences into a trie structure; and creating the index comprising:

generating a current sequential ID and a maximum sequential ID pair for each of the one or more trie nodes, wherein the current sequential ID of any descendant of a given trie node is between the current sequential ID of the given trie node and the maximum sequential ID;

generating an iso-depth link for each unique symbol in each of the one or more one-dimensional sequences, wherein the iso-depth link comprises trie nodes under the symbol; and

generating an offset list comprising an original position of each of the one or more subsequences in the weighted-sequence.

16. A machine-readable medium having instructions stored thereon for execution by a processor to perform a method of matching a query sequence in a weighted-sequences index, comprising the steps of:

receiving the query sequence;

transforming the query sequence into a weighted sequence;

encoding the weighted sequence into one or more one-dimensional sequences; and

searching one or more iso-depth links of the weighted-sequences index structure using the one or more one-dimensional sequences and returning an offset.